Computational Intelligence and Neuroscience 2016 vol.2016

Knowledge-Driven Event Extraction in Russian: Corpus-Based Linguistic Resources

Solovyev V., Ivanov V.

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2016 Valery Solovyev and Vladimir Ivanov.Automatic event extraction form text is an important step in knowledge acquisition and knowledge base population. Manual work in development of extraction system is indispensable either in corpus annotation or in vocabularies and pattern creation for a knowledge-based system. Recent works have been focused on adaptation of existing system (for extraction from English texts) to new domains. Event extraction in other languages was not studied due to the lack of resources and algorithms necessary for natural language processing. In this paper we define a set of linguistic resources that are necessary in development of a knowledge-based event extraction system in Russian: a vocabulary of subordination models, a vocabulary of event triggers, and a vocabulary of Frame Elements that are basic building blocks for semantic patterns. We propose a set of methods for creation of such vocabularies in Russian and other languages using Google Books NGram Corpus. The methods are evaluated in development of event extraction system for Russian.

http://dx.doi.org/10.1155/2016/4183760